

PMatch: Paired Masked Image Modeling for Dense Geometric Matching

==== Supplementary Material ====

Shengjie Zhu and Xiaoming Liu
 Department of Computer Science and Engineering,
 Michigan State University, East Lansing, MI, 48824
 zhusheng@msu.edu, liuxm@cse.msu.edu



Figure 1. **Visualization of produced co-planar points in MegaDepth Dataset.** The red point is one anchor point \mathbf{p}_m , while the green dots are the co-planar pixels among K sampled candidates $\{\mathbf{q}_n^m \mid 1 \leq n \leq K\}$, computed following Eqn. 1.

Production of the indicator matrix \mathcal{O}^+ . In the main paper Eqn. 14, we utilize an indicator matrix \mathcal{O}^+ to indicate the co-planar pairs between anchor and candidate points. Given the K anchor points $\{\mathbf{p}_m \mid 1 \leq m \leq K\}$ and $K \times K$ candidate points $\{\mathbf{q}_n^m \mid 1 \leq n \leq K\}$, the indicator matrix \mathcal{O}^+ of size $K \times K$ is computed as:

$$\mathcal{O}_{m,n}^+ = 1 \quad \text{if} \quad \begin{cases} 1 - \arccos(\mathbf{n}_{\mathbf{p}_m}^\top \mathbf{n}_{\mathbf{q}_n^m}) < k_1 \\ \text{dist}(\mathbf{n}_{\mathbf{p}_m}, \mathbf{p}_m, d_{\mathbf{p}_m}, \mathbf{q}_n^m, d_{\mathbf{q}_n^m}, \mathbf{K}_1) < k_2 \\ \|\text{proj}(\mathbf{H}_{\mathbf{p}_m}^\top, \mathbf{q}_n^m) - \mathbf{q}_n^m\|_2 < k_3. \end{cases} \quad (1)$$

The function $\text{dist}(\cdot)$ computes the point-to-plane distance in 3D space. The plane is spanned by the norm $\mathbf{n}_{\mathbf{p}_m}$ and the re-projected 3D point at \mathbf{p}_m . The function $\text{proj}(\cdot)$ indicates the planar projection under pixel \mathbf{p}_m homography matrix $\mathbf{H}_{\mathbf{p}_m}$ (see main paper equation Eqn. 13). The k_1 , k_2 and k_3 are set to 0.002, 0.02, and 1 respectively. We visualize the produced groundtruth co-planar points in Fig. 1.

Additional Visualization. We include additional visualization of the paired MIM pretexting task (Fig. 2) and reconstruction visualization (Fig. 4 and Fig. 3).

References

- [1] Zhengqi Li and Noah Snavely. Megadepth: Learning single-view depth prediction from internet photos. In *CVPR*, 2018.

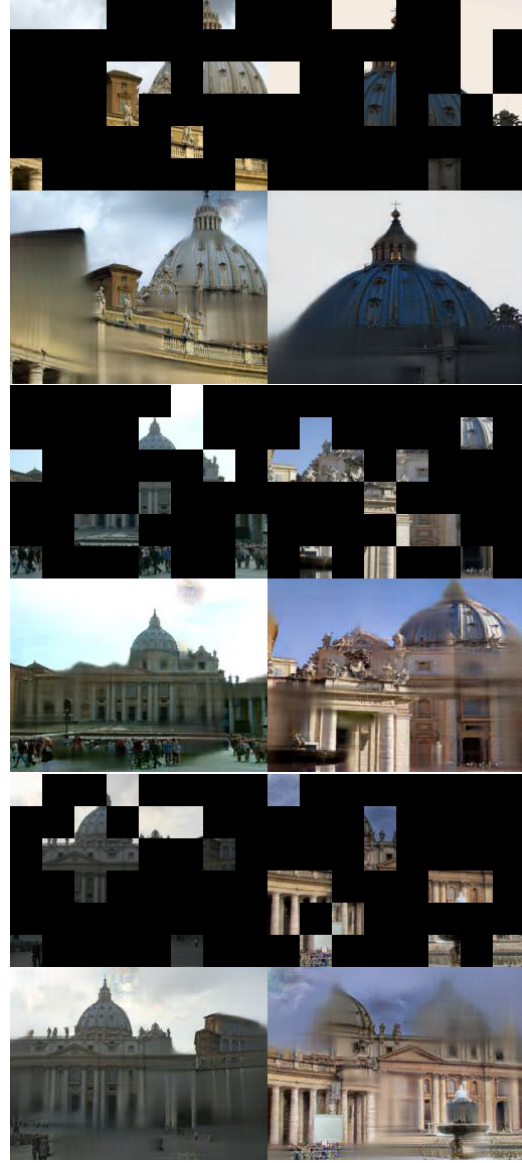


Figure 2. **Visual Quality of the paired MIM pretext task.** Visualized cases are from the MegaDepth and the ScanNet dataset.

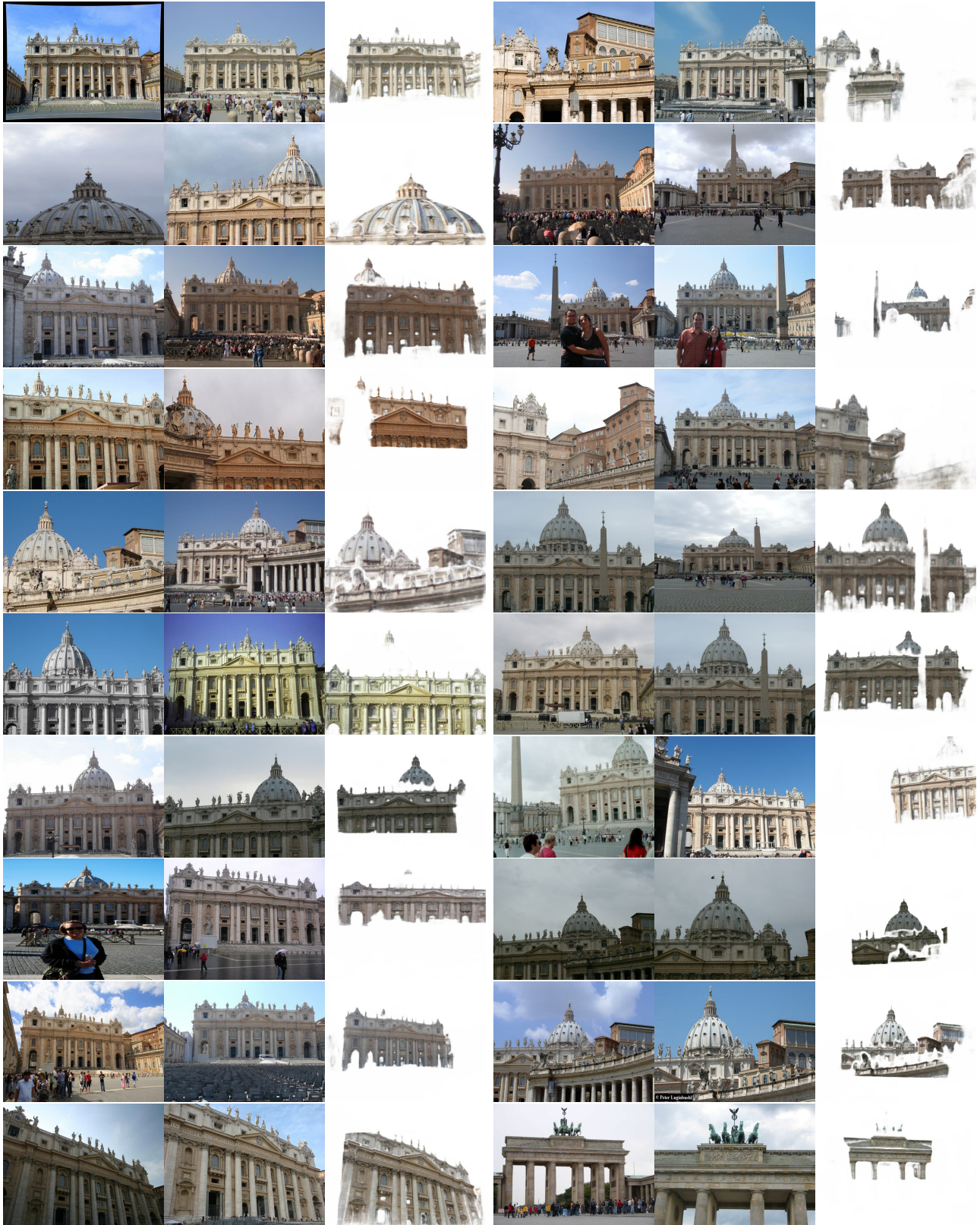


Figure 3. Visual Quality of the Reconstruction on MegaDepth [1].



Figure 4. Visual Quality of the Reconstruction on ScanNet.